# Feature extraction, feature-indexed databases, features in musicology and evolution with feature; Also, features

#### Cory McKay

Marianopolis College, Canada

CIRMMT, Canada

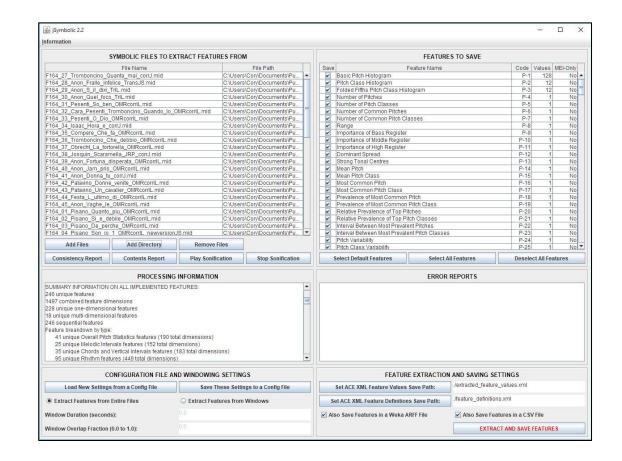
May 25, 2023.

CIRMMT Scientific Event

Montreal, Canada

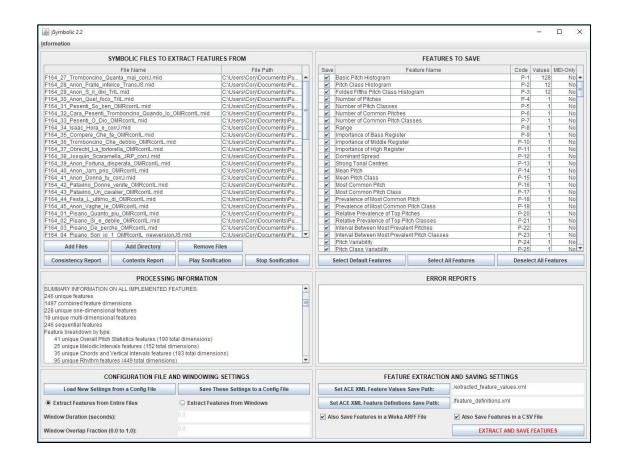
# Extracting features with jSymbolic

- The jSymbolic research software (2004 to present) automatically extracts features from symbolic digital scores (MIDI and MEI)
  - Open source
- Features used for:
  - Training models using machine learning
  - Direct musicological analysis
    - Statistical or manual



# jSymbolic's feature types

- Pitch statistics
  - e.g. Range
- Melody / horizontal intervals
  - e.g. Most Common Melodic Interval
- Chords / vertical intervals
  - e.g. Vertical Minor Third Prevalence
- Texture
  - e.g. Parallel Motion
- Rhythm
  - e.g. Note Density per Quarter Note
- Instrumentation
  - e.g. Note Prevalence of Unpitched Instruments
- Dynamics
  - e.g. Variation of Dynamics



# Extracting features with jSymbolic

- The newest version (jSymbolic3) is currently undergoing final code review and testing
- Extracts 533 unique features and 2040 feature values
  - Up from the current (2.2) release version's 246 and 1497, respectively
- Includes new features based on ngrams
  - Horizontal / melodic
  - Vertical / harmonic
  - Rhythmic

|  | SYMBOLIC FILES TO EX   | TRACT FEATUR  | ES FROM               | i Y                       |   |           | FEATURE                                    | S TO SAVE              |               |            |          |
|--|--|---|-----------------------|---------------------------|---|-----------|--|------------------------|---------------|------------|----------|
|  | File Name  |   | 1                     | File Path                 |   | ave       | Feature Name                               |                        | Code          | Values     | MEI-Only |
| F164 27 Tromboncino Q  |  |   | Cillingra             | Cor/Documents\Pu          |   |           | sic Pitch Histogram                        |                        | P-1           | 128        | No       |
| F164_28_Anon_Fralle_infe   |  |   |                       | Con/Documents/Pu          |   |           | ch Class Histogram                         |                        | P-2           | 120        | No       |
| F164_29_Anon_S_II_dixi_  |  |   |                       | Con/Documents/Pu          |   |           | Ided Fifths Pitch Class Histogram          |                        | P-3           | 12         | No       |
| F164 30 Anon Quel foco   |  |   |                       | Cory/Documents/Pu         |   |           | mber of Pitches                            |                        | P-4           | 1          |          |
| F164 31 Pesenti So ber   |  |   |                       | Con/Documents/Pu          |   |           | imber of Pitch Classes                     |                        | P-5           | 1          |          |
|  | romboncino Quando lo C   | MRcorrll mid  |                       | Con/Documents/Pu          |   |           | imber of Common Pitches                    |                        | P-6           | 1          |          |
| F164 33 Pesenti O Dio  |  | in toorna. The  |                       | Cory/Documents/Pu         |   |           | mber of Common Pitch Classes               |                        | P-7           | 1          |          |
| F164 34 Isaac Hora e c   |  |   |                       | Cor/Documents/Pu          |   |           | inge                                       |                        | P-8           | 1          |          |
| F164_35_Compere_Che_   |  |   |                       | Con/Documents\Pu          |   |           | portance of Bass Register                  |                        | P-9           | 1          |          |
|  | he_debbio_OMRcorrlL.mid  | d.  |                       | Con/Documents\Pu          |   |           | portance of Middle Register                |                        | P-10          | 1          |          |
| F164 37 Obrecht La tort  |  |   |                       | Cor/Documents\Pu          |   |           | portance of High Register                  |                        | P-11          | 1          |          |
| F164 38 Josquin Scaran   |  |   |                       | Con/Documents/Pu          |   |           | minant Spread                              |                        | P-12          | 1          |          |
| F164_39_Anon_Fortuna_c   |  |   |                       | Cor/Documents/Pu          |   |           | ong Tonal Centres                          |                        | P-13          | 1          |          |
| F164 40 Anon Jam pris  |  |   |                       | Cory/Documents/Pu         |   |           | an Pitch                                   |                        | P-14          | 1          |          |
| F164 41 Anon Donna tu  |  |   |                       | Con/Documents/Pu          |   |           | an Pitch Class                             |                        | P-15          | 1          |          |
| F164 42 Patavino Donne   |  |   |                       | Con/Documents/Pu          |   |           | st Common Pitch                            |                        | P-16          | 1          |          |
| F164 43 Patavino Un ca   |  |   |                       | Cor/Documents/Pu          |   |           | st Common Pitch Class                      |                        | P-17          | 1          |          |
| F164 44 Festa L ultimo   |  |   |                       | Cor/Documents/Pu          |   |           | evalence of Most Common Pitch              |                        | P-18          | 1          |          |
| F164_45_Anon_Vaghe_le  |  |   |                       | Con/Documents/Pu          |   |           | evalence of Most Common Pitch Class        |                        | P-19          | 1          |          |
| F164_01_Pisano_Quanto  |  |   |                       | Cory/Documents/Pu         |   |           | lative Prevalence of Top Pitches           |                        | P-20          | 1          |          |
| F164 02 Pisano Si e de   |  |   |                       | Cor/Documents/Pu          |   |           | lative Prevalence of Top Pitch Classes     |                        | P-21          | 1          |          |
| F164 03 Pisano De per  |  |   |                       | Con/Documents/Pu          |   |           | erval Between Most Prevalent Pitches       |                        | P-22          | 1          |          |
|  | 1 OMRcorril newversion.  | IS mid  |                       |                           |   |           | erval Between Most Prevalent Pitch Classes | 2                      | P-23          | 1          |          |
| 1104 04 HISBIR SOIL 10   | T OMITCOME NEWTOISIONS   | U.IIIIG   | 10.103013             | ioonioocumentoir u j      |   |           | ch Variability                             | ,                      | P-24          | 1          |          |
| Add Files  | Add Directory  | Remove F  | iles                  |                           |   |           | ch Class Variability                       |                        | P-25          | 1          | No       |
| Consistency Report   | Contents Report  | Play Sonific  | ation                 | Stop Sonification         |   | 1000 0000 |  | II Features            | Desele        | ct All Fei | 10.000   |
|  |  | INFORMATION   |                       |                           |   |           | ERROR                                      | REPORTS                |               |            |          |
| 246 unique features  |  |   |                       |                           |   |           |  |                        |               |            |          |
| 25 unique Melodic Inte<br>35 unique Chords and   | nal features   | nensions)<br>183 total dimensio                           | ons)                  |                           | × |           |  |                        |               |            |          |
| 228 unique one-dimensior<br>18 unique multi-dimensior<br>246 sequential features<br>Feature breakdown by type<br>41 unique Overall Pitch<br>25 unique Melodic Inte<br>35 unique Chords and<br>95 unique Rhythm feat                          | nal features<br>:<br>statistics features (190 tot:<br>rvals features (152 total dim<br>Vertical Intervals features (1  | nensions)<br>183 total dimensio<br>)                      |                       | s                         | - |           | FEATURE EXTRACTIO                          | N AND SAVING SETTI     | VGS           |            |          |
| 228 unique one-dimension<br>18 unique multi-dimension<br>246 sequential features<br>Feature breakdown by type<br>41 unique Overall Pitch<br>25 unique Melodic Inte<br>35 unique Chords and<br>95 unique Rhythm feat                          | nal features<br>Statistics features (190 tot<br>rvals features (152 total dim<br>Vertical Intervals features (*<br><u>ures (449 total dimensions</u> )   | nensions)<br>183 total dimensio<br>)<br>ID WINDOWING      | SETTING               | :S<br>gs to a Config File |   | Set       | FEATURE EXTRACTION                         | AND SAVING SETTI       |               |            |          |
| 228 unique one-dimension<br>18 unique multi-dimension<br>48 sequential features<br>Feature breakdown by type<br>41 unique Overall Pikt<br>25 unique Melodic Inte<br>35 unique Chords and<br>95 unique Rhythm feat<br>CC<br>Load New Settings | al features<br>Statistics features (190 tot:<br>vertical intervals features (<br>vertical intervals features (<br>ures (449 total dimensions)<br>ONFIGURATION FILE AN<br>s from a Config File                | nensions)<br>183 total dimensio<br>)<br>ID WINDOWING      | SETTING<br>ese Settin | gs to a Config File       |   |           |  |                        | ues.xml       |            |          |
| 228 unique one-dimension<br>18 unique multi-dimension<br>246 sequential features<br>Feature breakdown by type<br>41 unique overall Pitch<br>25 unique Alelodic Inte<br>35 unique Chords and<br>95 unique Rhythm feat                         | tal features<br>Statistics features (190 tot<br>vals features (192 total dim<br>Vertical Intervals features (<br>ures (449 total dimensions)<br>ONFIGURATION FILE AN<br>6 from a Config File<br>Entire Files | nensions)<br>183 total dimensio<br>D WINDOWING<br>Save Th | SETTING<br>ese Settin | gs to a Config File       |   | Set A     | ACE XML Feature Values Save Path:          | /extracted_feature_val | ues.xml<br>nl | / File     |          |

May 25, 2023.

Current musicological research projects based on features (with CIRMMT partners)

- Origins of the madrigal
  - Cumming & McKay 2018; Cumming & McKay 2021
  - Soon to be revisited with an expanded corpus
- Josquin attribution certainty
  - McKay, Cumming & Fujinaga 2017
  - Currently working on new a book chapter validating published source-based judgements with content-based feature analysis
- Empirically delineating the styles of Josquin and Pierre de la Rue
  - Cumming, McKay, Schubert, Nápoles López & Margot (in press)

# Current musicological research projects based on features (with Iberian partners)

- Regional style in Iberian Renaissance music
  - Musical influences of Pedro Fernández Buch (Cuenca & McKay 2022)
  - Musical Influences of Cristóbal de Morales and Francisco Guerrero (McKay & Cuenca 2021)
- Attribution of anonymous and doubtfully attributed works
  - Masses transcribed by Siro Cisilino (Cuenca & McKay 2023)
  - Coimbra manuscripts (Cuenca & McKay 2021)
  - Ave verum corpus and O decus virgineum (Rodríguez-García & McKay 2021)
  - Ave festiva ferculis (Rodríguez-García & McKay 2021)
  - Gaffurius Codices (McKay 2021)
- Currently part of an application for a Proyectos de Generación de Conocimiento grant in Spain

### SIMSSA DB

- A prototype database of symbolic music files intended for research in computational musicology
  - And, eventually, associated images, audio recordings, texts, etc.
- Searchable by:
  - Free text
  - Faceted metadata
  - Feature values
    - Auto-annotated by jSymbolic

| Search                | 3 Musical Works for query "Thomme armé" and selected facets | Please note that features only  |  |
|-----------------------|---|---|--|
| Thomme armé           |   | apply to valid MIDI, Music XML  |  |
| Sort By               | к 1 ж   | and MEI files, and will exclude file<br>formats from Sibelius, Finale, etc. |  |
| Eest Match            | Missa L'homme armé l  | For an explanation of all features,   |  |
|                       | Composer(s): La Rue, Pierre 14601518                        | please consult the jSymbolic<br>Manual.                                     |  |
| Composition Year From |   |   |  |
|                       |   | Chords and Vertical Interval<br>Features                                    |  |
| Composition Year To   | Missa L'homme armé super voces musicales                    | reduies   |  |
|                       | Composer(s): des Prez, Josquin 14401521                     | Dynamics Features   |  |
| Genre (Type of Work)  | Section(s):   | Instrumentation Features  |  |
| Mass(3)               | * Sanctus (Benedictus)                                      |   |  |
|                       | * Sanctus (In nomine)                                       | Melodic Interval Features   |  |
| Genre (Style)         |   | Amount of Arpeggiation:<br>0.1389 - 0.36105600000000                        |  |
| Renaissance(3)        | Sanctus (Qui venit)   | 0.1303 - 0.3010300000000  |  |
|                       | <ul> <li>Genres (Type of Work): Mass</li> </ul>             | Average Interval Spanned by<br>Melodic Arcs;                                |  |
| Composer              |   | 3.746 - 6.556   |  |
| La Rue, Pierre(1)     | <ul> <li>Genres (Style): Renaissance</li> </ul>             |   |  |
|                       | File(s) Holding Complete Musical Work:                      | Average Length of Melodic Arcs:   |  |
| des Prez, Josquin(2)  | <ul> <li>File(s) Holding an Individual Section:</li> </ul>  | 1.444 - 3.667   |  |
|                       | Sanctus (Benedictus)  | Chromatic Motion:   |  |
| Instrument/Voice      | sancius (penedicius)  | 0.1556 - 0.33   |  |
| Sacred or Secular     | km  | Direction of Melodic Motion:  |  |
| Sacred(3)             |   | 0.4255 - 0.5909   |  |
|                       |   | Distance Between Most Prevalent   |  |
| File Format           | .pdf 🔹  | Melodic Intervals:  |  |
| □ .mid(9)             |   | 1+1 C   |  |
|                       |   | Mean Melodic Interval:  |  |
| 5. 50m(2)             | .mid 🔸  | 1.917 - 2.688   |  |
| .pdf(9)               |   | Melodic Embellishments:   |  |
|                       |   | Melodic Embellishments:<br>0 - 0.05556                                      |  |
| □ km/9)               | · · · · · · · · · · · · · · · · · · ·                       |   |  |

May 25, 2023.

### SIMSSA DB

- Additional priorities:
  - Chains of provenance
  - Authority control (using VIAF)
  - Models relationships between data, sources and (abstract) works, sections and parts
  - Corpora can be associated with specific studies (e.g., for repeatability of research)
- To be linked with other music repositories as part of the long-term LinkedMusic project (2022-2029)
  - Involves other CIRMMT members, including Ichiro Fujinaga, Julie Cumming and Audrey Laplante, as well as numerous international partners

| Search                | 3 Musical Works for guery "I'homme armé" and selected facets | Please note that features only  |
|-----------------------|--|---|
| Thomme armé           |  | apply to valid MIDL Music XML   |
|                       | « 1 »  | and MEI files, and will exclude file<br>formats from Sibelius, Finale, etc. |
| Sort By               |  | For an explanation of all features,   |
| Best Match            | Missa L'homme armé l   | please consult the jSymbolic  |
| Composition Year From | Composer(s): La Rue, Pierre 14601518                         | Manual  |
|                       |  | Chords and Vertical Interval  |
| Commercial on Mana Ta |  | Features  |
| Composition Year To   | Missa L'homme armé super voces musicales                     | Description of the  |
|                       | Composer(s): des Prez, Josquin 14401521                      | Dynamics Features   |
| Genre (Type of Work)  | Section(s):  | Instrumentation Features  |
| Mass(3)               | Sanctus (Benedictus)   | Melodic Interval Features   |
|                       | * Sanctus (In nomine)  |   |
| Genre (Style)         |  | Amount of Arpeggiation:<br>0.1389 - 0.36105600000000                        |
| Renaissance(3)        | Sanctus (Qui venit)  | 0.1389 - 0.36105600000001   |
|                       | <ul> <li>Genres (Type of Work): Mass</li> </ul>              | Average Interval Spanned by   |
| Composer              | Genres (Type of Work): Mass                                  | Melodic Arcs:<br>3.746 - 6.556  |
| La Rue, Pierre(1)     | <ul> <li>Genres (Style): Renaissance</li> </ul>              | 3.746 - 6.556   |
| La rue, Pierre(I)     | File(s) Holding Complete Musical Work:                       | Average Length of Melodic Arcs:   |
| des Prez, Josquin(2)  | rinely rooming complete matrice more.                        | 1.444 - 3.667   |
|                       | <ul> <li>File(s) Holding an Individual Section:</li> </ul>   | Chromatic Motion:   |
| Instrument/Voice      | Sanctus (Benedictus)   | 0.1556 - 0.33   |
|                       |  | 0.1550 - 0.55   |
| Sacred or Secular     | .km +  | Direction of Melodic Motion:  |
| Sacred(3)             |  | 0.4255 - 0.5909   |
| File Format           |  | Distance Between Most Prevalent   |
|                       | .pdf   | Melodic Intervals:  |
| .mid(9)               |  | 1-1 0   |
|                       | .mid 🗖   | Mean Melodic Interval:  |
|                       |  | 1.917 - 2.688   |
| .pdf(9)               |  | Melodic Embellishments:   |
|                       |  | 0 - 0.05556   |
| □ km(9)               | • .sto •   | v - 0.05556   |
| Filter                |  | Melodic Large Intervals:  |

May 25, 2023.

### Multi-modal music classification

- My original doctoral dissertation (2010) focused on multi-modal music classification
  - Investigating if, when and how information drawn from different types of musical data can be usefully combined using machine learning
  - jSymbolic was originally just part of this project
- Currently working on using genetic algorithms to explore different combinations of modalities and feature types
  - Features extracted from audio, lyric texts, symbolic scores, album cover images, semantic tags and playlist co-occurrences
  - Evaluated based on "importance," "redundancy" and "stability"
  - Vatolkin & McKay 2022a; Vatolkin & McKay 2022b
- Currently part of an application for DFG funding in Germany

# Thanks for your attention

cory.mckay@mail.mcgill.ca



Social Sciences and Humanities Research Council of Canada Conseil de recherches en sciences humaines du Canada

Canada









Centre for Interdisciplinary Research in Music Media and Technology